



# Kepler Cohort 2014 Evaluation Endline Report

December 2015

## Contents

Executive Summary .....	3
Introduction .....	4
Objective of Year 2 evaluation.....	4
Methodology.....	5
Design .....	5
Selection process for 2014 cohort comparison group.....	6
Baseline characteristic averages between Kepler and comparison groups .....	6
Metrics .....	7
Test Results .....	8
Presentation of results .....	8
Summary of results.....	9
Critical thinking .....	10
Cognitive skills .....	10
English language .....	10
Computer literacy .....	10
Survey Findings .....	11
Internships .....	11
Perceived job prospects.....	14
Housing .....	17
Academic dishonesty.....	18
Communication with Kepler administration.....	19
Summer plans.....	20
Conclusions.....	22
Appendix A: Technical details on methodology .....	23
Appendix B: Sample test questions .....	27
Appendix C: Year 2 endline raw test score statistics .....	32
Appendix D: Regression results.....	33
Appendix E: Robustness checks on effect of English skills on critical thinking test scores ....	34
Appendix F: Subgroup results .....	37

## Executive Summary

Kepler, a university program in Rwanda which launched in 2013, aims to develop a high-quality, low-cost model of higher education. The program provides a blended learning model with Massive Open Online Courses (MOOCs)<sup>1</sup> and in-person instruction with a focus on improving job market outcomes for students. IDinsight partnered with Kepler during the 2014-2015 academic year (“Year 2”) to evaluate the impact of the Kepler program during their second year of operation relative to traditional Rwandan universities on students’

1. Critical thinking,
2. Cognitive skills (including English grammar and vocabulary, math, and logic),
3. English language skills, and
4. Computer literacy.

Results of the evaluation are as follows.<sup>2</sup> All impacts are given as standard effect sizes, calculated as the difference between the Kepler students’ and comparison students’ average test scores in terms of standard deviations (SD) of comparison students’ scores. This gives a sense of how large differences in test scores are relative to the spread in scores and allows for easy comparison across different tests. Results are also expressed as percentage point (pp) differences in test scores calculated as percentage of total possible points on the test:

1. **Critical thinking: 0.53 SD (4.9 pp)**, as measured by the Watson-Glaser critical thinking test<sup>3</sup>
2. **Cognitive skills: English: 0.45 SD (9.0 pp), math: 0.14 SD (2.5 pp, not statistically significant), logic: 0.41 SD (7.9 pp)**
3. **English: Reading: 0.39 SD (5.8 pp), writing: 0.60 SD (2.8 pp)**, as measured by the International English Language Testing System (IELTS) reading and writing tests)
4. **Computer literacy: 1.90 SD (26.5 pp)**, as measured by a test that included typing speed, web research, assessing website credibility, Microsoft Word, Microsoft Excel, and email

As indicated by the standard effect sizes, **all of the statistically significant results above are particularly large in magnitude**: in education studies, effect sizes larger than 0.3 SD are considered large, while effect sizes larger than 0.5 SD are considered very large.<sup>4</sup>

Kepler and comparison students of both cohorts responded to survey questions on various aspects of their university experience. Survey findings related to preparation for the job market show that more Kepler students than comparison students are engaging in internships earlier on in their university education, and are receiving more feedback on their internships. While Kepler students perceive their university to be performing mostly at a similar level as comparison students, though less strong in reputation and importance of coursework, they feel that they have more personal connections they can use in finding jobs and are more confident in their likelihood to find jobs after graduation.

---

<sup>1</sup> MOOCs are comprised of online course modules that students can access and view at their own pace.

<sup>2</sup> All results are statistically significant at the 5% significance level unless otherwise noted.

<sup>3</sup> Robustness checks were performed to see how much improvements in English skills can account for improvements in Watson-Glaser test scores for the first year impact for the 2014 cohort. Results are somewhat unclear, but suggest that English skills may account for some of the increase in Watson-Glaser scores.

<sup>4</sup> J-PAL Policy Lessons on Education – Student Learning, Background. <http://www.povertyactionlab.org/policy-lessons/education/student-learning?tab=tab-background>. Accessed July 2015.

## Introduction

Kepler, a university program in Rwanda which launched in 2013, aims to develop a high-quality, low-cost model of higher education. The program provides a blended learning model with Massive Open Online Courses (MOOCs) and in-person instruction with a focus on improving job market outcomes for students. The Kepler program stresses critical thinking skills, English comprehension, and computer literacy. Kepler started with its inaugural cohort of 50 students in 2013 and its second cohort of approximately 100 students in 2014.<sup>5</sup>

Tests were administered for the 2014 cohort in June and September/October 2014 (“Year 2 baseline” for Kepler and comparison students, respectively) and in April/May 2015 (“Year 2 endline”).

Study results are presented in the following sections: *a)* Introduction, *b)* Methodology, *c)* Test Results, *d)* Survey Findings, and *e)* Conclusions.

## Objective of Year 2 evaluation

The **objective for the Year 2 evaluation** is to measure the first year impact of the Kepler program for the 2014 cohort. IDinsight partnered with Kepler during the 2014-2015 academic year (“Year 2”) to evaluate the impact of their program. The Year 2 evaluation includes an evaluation of the quantitative impact of attending the Kepler program on students in the 2014 cohort, as well as survey results from both the 2013 and 2014 cohorts. This evaluation was able to capture baseline metrics for Kepler students and comparison students prior to their receiving any substantial university programming.<sup>6</sup> This allows us to ensure that the Kepler and comparison students were similar in learning outcome levels before starting university. Incorporating baseline scores in the impact measurement also increases statistical precision when estimating differences in endline scores.

The 2014 cohort’s first year experience at Kepler differs from that of the 2013 cohort in several ways, including the introduction of an extensive bridge program that aims to prime students’ English language, computer literacy, and other skills important for success in the Kepler program; having a greater number of students in the cohort; and having upperclass peers who could serve as peer mentors. In addition, the recruitment process was modified, yielding a class with slightly higher skill levels in some areas. At the same time as changes were taking place for incoming students, the second year students enter a phase where learning is more self-directed than the foundational first year.

---

<sup>5</sup> Several students dropped out of the Kepler program after the start of the program, with 88 students remaining in the 2014 cohort as of April 2015. Most students left Kepler to attend the University of Rwanda on scholarships, mostly to study sciences, or for scholarships abroad, which they found out about after starting at Kepler.

<sup>6</sup> The baseline tests were administered in the first week of the bridge program for Kepler students, and sometime between pre-registration and the third week of school for comparison students. Regressions were run of baseline scores on the date of baseline exam for comparison students; no significant positive correlations between a later test date and test score were found systematically across tests, suggesting that the first few weeks of school did not have a large impact on learning outcomes for comparison students.

## Methodology

### Design

As with all impact evaluations, the objective of the evaluator is to create a comparison group that mimics as closely as possible how the treatment group—in this case, the Kepler students—would perform in the absence of treatment—that is, if they had not attended Kepler. While randomly assigning treatment—such as invitations to attend Kepler—would be the most rigorous way of creating similar comparison and treatment groups, this was not possible.<sup>7</sup> Instead, IDinsight used the most rigorous design possible given the circumstances: a matching methodology which matches each Kepler student to one or two students<sup>8</sup> at traditional Rwandan universities on a variety of observed characteristics that correlate with learning outcomes.

This study used a propensity score matching methodology to identify non-Kepler students who would have had a high likelihood of enrolling in Kepler, given the opportunity, based on key characteristics such as prior education, socio-economic status, and baseline test scores.<sup>9,10</sup> This process constructs a group of comparison students who are statistically similar to the Kepler students in terms of these key measurable characteristics. The similarity of the groups increases our confidence that differences in endline test scores between Kepler and comparison students can be attributed to the Kepler program itself and not to other confounding factors.<sup>11</sup>

The first year impact of the Kepler model for the 2014 cohort was estimated by calculating the difference in the average 2014 cohort's endline test results between Kepler and comparison students. All estimates controlled for important covariates<sup>12</sup> such as student age, gender, previous academic achievement, family background, and other relevant characteristics.<sup>13</sup> Year 2 baseline scores from the same test were also included as covariates to adjust for prior learning outcomes for each test result.

---

<sup>7</sup> While Kepler was open to the idea of randomizing admissions for a portion of their incoming class, ultimately their admissions process did not identify enough qualified students to be able to randomize without sacrificing either size or quality of the incoming student cohort.

<sup>8</sup> For the 2014 cohort, students were first stratified by performance on certain baseline tests. In strata where the ratio of comparison to Kepler students was greater than 2:1, two comparison students were matched to every one Kepler student; in strata where the ratio was smaller, some Kepler students were only matched to one comparison student. Please refer to Appendix A for further details on the stratification used in matching.

<sup>9</sup> The non-Kepler students included in this study all had not heard of Kepler. Because the reach of Kepler's recruiting is still limited, it is likely that these students had not heard of Kepler for arbitrary reasons, and may have been likely to enroll had they known of Kepler. Several students originally identified for the comparison group were extended offers to enroll at Kepler, and most (five students) accepted, supporting this claim.

<sup>10</sup> Please refer to Appendix A for further details on the propensity score matching methodology.

<sup>11</sup> Matching on observable characteristics does not guarantee that there is balance on unobservable characteristics, such as motivation. It is possible that differences in unobservable characteristics may have some impact on results. Please see Appendix A for further discussion on limitations of matching.

<sup>12</sup> Covariates refer to characteristics that are likely to affect the outcome of interest.

<sup>13</sup> Please see Appendix A for a full list of covariates used.

Further details on the matching methodology used, regression analysis to estimate impacts, and discussion on limitations of the study are included in Appendix A.

Quantitative results from the survey questions are presented as descriptive statistics reflecting averages of answers or percentage of students responding a certain way. Key points are also summarized for responses to open-ended questions.

### Selection process for 2014 cohort comparison group

For the 2014 cohort, 143 comparison students were selected for the Year 2 endline tests from an initial pool of 2,868 students attending five different universities.<sup>14</sup> The selection process was designed to be similar to the Kepler admissions process in order to identify students who were most similar to the Kepler students. The following figure outlines the steps taken to select the comparison group:



Candidates were initially recruited from comparison university registrars and pre-term registration activities and invited for further screening based on their Senior 5 and Senior 6 marks. Any students who had applied to Kepler or had taken the national secondary school exam prior to 2011 were excluded. Students were then screened with the same tests and interview processes that were used in Kepler admissions, including an English test, math test, one-on-one oral exam to screen for English speaking fluency, and one-on-one interviews.<sup>15</sup> Based on scores on these tests, students were invited to take part in baseline testing, during which they took the baseline learning outcome tests as well as a survey collecting information on their personal and household characteristics. Baseline data (including test scores and personal characteristics) were used to match 143 comparison students with the 88 Kepler 2014 cohort students. These matched students were then invited to take part in the Year 2 endline testing.

### Baseline characteristic averages between Kepler and comparison groups

For the 2014 cohort, a greater proportion of Kepler students attended private secondary schools,<sup>16</sup> though Kepler and comparison students scored very closely on both Senior 5 and Senior 6 exams. Kepler students reported somewhat higher expected post-graduation

---

<sup>14</sup> Students were sampled from the following five universities in Kigali: Adventist University of Central Africa; Kigali Independent University; and the following University of Rwanda Colleges: College of Business and Economics, College of Education, and the College of Science and Technology. While all students were initially screened in Kigali, a 2 students from the 2014 cohort shifted to study at the Huye campus in southern Rwanda.

<sup>15</sup> Initially comparison students also took part in a 50 minute group problem solving and discussion exercise, but results of this exercise had low correlation to admission outside of the other exams, and was therefore dropped.

<sup>16</sup> While the type of secondary school (public/private) was not included as a covariate in analysis for the results presented here, analysis was re-run including it as a covariate to see if the imbalance between Kepler and comparison students drives any of the results obtained. No results substantially changed in magnitude or degree of statistical significance.

earnings but lower expected earnings five years after graduation, with the latter difference not being statistically significant. There are no statistically significant differences in other household or demographic characteristics or in any of the baseline test scores, suggesting that the Kepler and comparison students are similar on average, particularly in learning outcomes prior to receiving any university education.

Please see Appendix A for the mean values of all covariates for Kepler and comparison students, as well as the difference in means (and  $p$ -values) between Kepler and comparison students of the 2014 cohort.

## Metrics

During the launch of the Kepler program, Kepler identified critical thinking, English language skills, and computer literacy, as among the main outcomes of interest in determining the success of the program. After a discussion about the prevalence of cognitive skills tests in measuring learning, this outcome was added. Each of the following tests was administered at the Year 2 baseline in June and October 2014 for the Kepler and comparison students respectively,<sup>17</sup> and at the Year 2 endline in April-May 2015. Sample questions for each test are included in Appendix B.

### *Critical thinking*

The Watson-Glaser II Critical Thinking Appraisal (“Watson-Glaser”) was administered to measure students’ critical thinking ability. The Watson-Glaser measures three key features of critical thinking, including: *a*) recognizing assumptions, *b*) evaluating arguments, and *c*) drawing conclusions. Students were given 50 minutes to complete the paper-based 40-question multiple-choice test.

### *Cognitive skills*

IDinsight designed and administered a multiple-choice cognitive skills test, which included questions on English grammar and vocabulary, math, and logic. The test was designed to be similar to popular standardized exams, with the design influenced by the Scholastic Level Exam (SLE) format and question ideas pulled from the SLE, SAT, ACT, and other practice test questions.<sup>18</sup> Students were given 20 minutes to complete the 30-question test.

### *English language*

IDinsight administered portions of the International English Language Testing System (IELTS) reading and writing sections. Questions were drawn from both General Training and Academic IELTS practice tests.

---

<sup>17</sup> Year 2 baseline testing was administered for the Kepler students in June 2014, prior to the start of their bridge program, and for comparison students in October 2014, either prior to or up to three weeks after the beginning of their university programs. The difference in timing is due to the different start dates of the academic year for Kepler and comparison universities.

<sup>18</sup> Intellectual property was respected in the process, and all test content was original, or else derived from public domain sources.

The IELTS reading test requires students to answer multiple-choice and fill-in-the-blank questions about several reading passages. Students had 40 minutes to answer 25 questions based on two passages for the Year 2 baseline and 28 questions based on three passages for the Year 2 endline.<sup>19</sup>

For the IELTS writing test, students had 40 minutes to write a short essay defending their own opinion on a general question presented in a brief essay prompt. Essays were graded on task achievement, coherence and cohesion, lexical resource, and grammatical range and accuracy.

### *Computer literacy*

IDinsight designed and administered a computer literacy test comprised of six different sections, including typing speed, internet research, assessing website credibility, Microsoft Word, Microsoft Excel, and email.<sup>20</sup> Each section was timed separately, with the entire computer literacy test lasting approximately 50 minutes. The aggregate computer literacy score is constructed by equally weighting each section's score expressed as percent of total points scored.<sup>21</sup> While there is technically no maximum typing speed, the maximum score for the typing speed test was set to 50 words per minute, the target typing speed for Kepler.

## Test Results

### Presentation of results

The following sections present quantitative results for the Year 2 evaluation. Differences in scores are expressed as standard effect sizes, calculated as the difference in the number of standard deviations (SD)<sup>22</sup> between the average score of the Kepler students and the average score of the comparison students. The standard effect size gives a sense of how much test scores differ relative to the spread in scores. In education studies, an effect size of less than 0.1 SD is typically considered small, while an effect size of more than 0.3 SD is considered large, and an effect size of more than 0.5 SD is considered very large.<sup>23</sup> Results are also expressed as percentage point (pp) differences in test scores calculated as percentage of total possible points on the test.

All test score averages and impact estimates presented in these results account for differences in covariates, including student demographics and baseline scores. As such, the average scores presented here may vary slightly from actual test scores that do not adjust for

---

<sup>19</sup> The recommended completion time for both exams was equivalent.

<sup>20</sup> Email was not included in the Year 2 baseline computer literacy test because of internet problems at the testing facility.

<sup>21</sup> Due to different sections of the computer literacy test being successfully administered at different testing periods, the aggregate computer literacy scores reported for each part of the analysis may include different sections of the computer literacy test. For all reported aggregate computer literacy scores, each included component is weighted equally.

<sup>22</sup> Standard deviations of the comparison students' Year 2 endline test scores were chosen as the unit to express standard effect sizes as they give a sense of the spread in test scores that would have been seen at the Year 2 endline in students who did not attend Kepler.

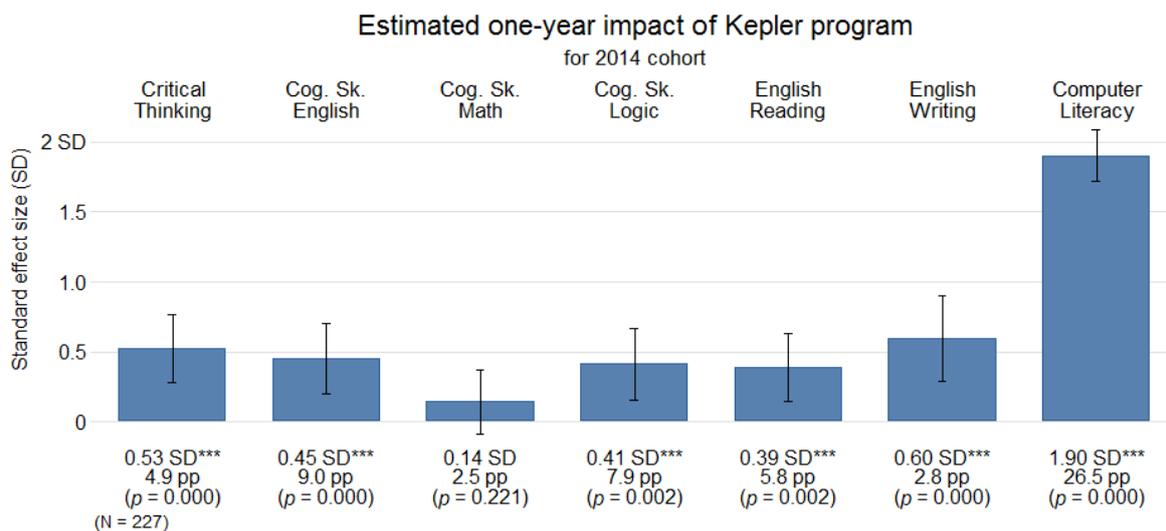
<sup>23</sup> See citation in Footnote 4.

differences in covariates. Please see Appendix C for unadjusted average scores. For full regression analysis table data, please refer to Appendix D.

## Summary of results

**Evaluation results indicate that Kepler students performed statistically significantly better than comparison students on nearly all Year 2 endline tests.** These scores reflect the effect of going to Kepler on these skill areas for the first year of the Kepler program for the 2014 cohort (Figure 1). The largest impact on student learning was in computer literacy skills by a sizeable margin, followed by English writing skills.

Figure 1



Bars depict the estimated one-year effect of the Kepler program on the indicated test scores, calculated as the difference in mean scores on Year 2 endline tests for Kepler and comparison students, adjusted for covariates. The estimates are also written beneath each bar in terms of standard deviations (SD) and percentage points (pp). Error bars depict the 95% confidence intervals. The  $p$ -values and asterisks correspond to tests of statistical significance, with lower  $p$ -values and more asterisks indicate greater confidence in the estimated impact and hence greater statistical significance of the result.<sup>24, 25</sup>

Subgroup analyses were conducted for gender and urban/rural domicile. Results from these subgroup analyses do not suggest that students at Kepler in these subgroups are improving at different rates relative to the students of the same subgroups at other universities; that is, the Kepler program does not appear to reduce either the gender gap or the urban/rural gap in academic achievement. Further details on subgroup analysis results are included in Appendix F.

<sup>24</sup> The  $p$ -value denotes the statistical significance of a test comparing averages: in this case, comparing average test scores between Kepler and comparison students. Using a 5% significance level, a low  $p$ -value (less than or equal to 0.05) means that it is unlikely that the observed difference between Kepler and comparison students' average test scores would have been observed if there were actually no impact—that is, the observed difference is *statistically significant*. A *high*  $p$ -value (greater than 0.05) means that the difference had a reasonable likelihood of occurring if there was actually no impact, and the difference between the two groups is *not statistically significant*.

<sup>25</sup> In this figure, one asterisk indicates a statistical difference at the 10% significance level, two asterisks indicate a statistical difference at the 5% significance level, and three asterisks represent a statistical difference at the 1% significance level.

## Critical thinking

**The 2014 Kepler cohort shows gains in critical thinking after their first year**, with a 0.53 SD (4.9 pp) difference in scores between Kepler and comparison students, a statistically significant result (Figure 1).

Note that the Watson-Glaser test does rely on English comprehension skills. Thus, it is possible that some of the difference is due to improved English language ability or to familiarity with taking multiple choice tests of this type. Robustness checks controlling for English scores were performed to determine how much improvements in English grammar and vocabulary scores can account for increases in Watson-Glaser scores for the first year impact for the 2014 cohort. While results may suggest that stronger English grammar and vocabulary skills could account for some of the increase in Watson-Glaser scores for Kepler students, these results were not conclusive. Please see Appendix E for more details. The clearest way to better isolate critical thinking skills from English skills in future evaluations may be to administer tests in Kinyarwanda.

## Cognitive skills

**The 2014 Kepler cohort performed better than comparison students on cognitive skills**, outperforming their peers on English grammar and vocabulary by 0.45 SD (9.0 pp, statistically significant), slightly better on math by 0.14 SD (2.5 pp, not statistically significant), and better on logic by 0.41 SD (7.9 pp, statistically significant) (Figure 1).

## English language

**The 2014 Kepler cohort performed better than the comparison students in English reading and writing at the end of their first year**, scoring higher on the IELTS reading exam by 0.39 SD (5.8 pp) and on the IELTS writing exam by 0.60 SD (2.8 pp),<sup>26</sup> both statistically significant results (Figure 1).

These results are in line with initial hypotheses that the Kepler program would serve to strengthen English skills, given that the Kepler program emphasizes using English in class much more heavily than traditional Rwandan universities.

## Computer literacy

**The 2014 Kepler cohort, after one year of Kepler programming, greatly outperformed comparison students on the computer literacy test** at the Year 2 endline (including all components), with an estimated difference of 1.90 SD (26.5 pp), almost doubling comparison students' adjusted mean score (Figure 1). This result suggests that the 2014 cohort acquired significant computer literacy skills in their first year.

---

<sup>26</sup> Although the effect on the IELTS writing exam may seem small in terms of percentage points, the large standardized effect size indicates that the effect is in actuality large when considering the narrow spread in scores.

## Survey Findings

In addition to the Year 2 endline tests, Kepler and comparison students from both 2013 and 2014 cohorts took an online survey asking about various aspects of their university experience: internships, students' perceived job prospects, potential job resources, housing while in university, academic dishonesty, communication with Kepler administration (for Kepler students), and plans for the summer and next school year.

### Internships

With the exception of the first question asking whether a student had an internship while in university, which was asked of all students, all other questions on internships were asked only of students who reported having had at least one internship. Students who had two internships answered questions about internships separately for each of their two internships; the two responses from these students are counted as independent answers in the findings reported below (such that we are looking at the percent of internships for which students reported certain answers, rather than the percent of students who responded with certain answers). For Kepler students, this includes 61 students reporting on 80 internships; for comparison students, this includes 33 students reporting on 37 internships.

Figure 2

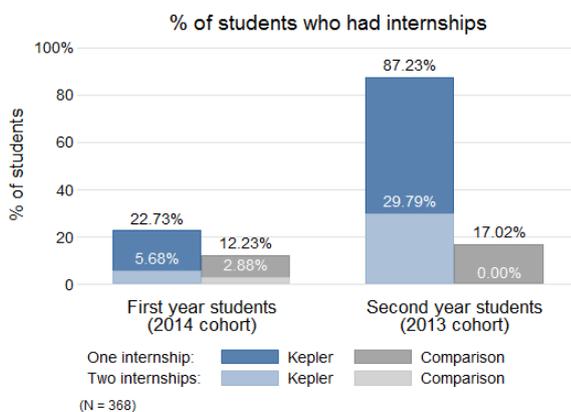
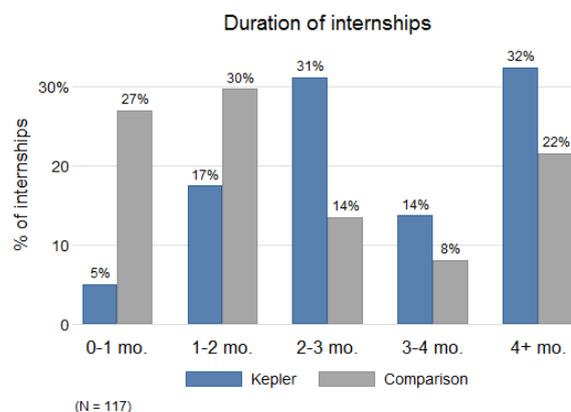


Figure 3



Note that only responses from students who reported having at least one internship are included. Students who reported having two internships answered this question for both of their internships, with each answer counting as being from a separate internship.

**Kepler students tend to have more and longer internships than comparison students, mostly during classes.** Kepler students in general tend to have had more internships than comparison students, particularly for second year students (2013 cohort) where nearly all Kepler students reported having had at least one internship while being a Kepler student (Figure 2).<sup>27</sup> (Note that this survey was taken before the summer following the first year at Kepler for the 2014 cohort, and many of the 2014 Kepler cohort may be taking on internships during the summer.)

<sup>27</sup> Speaking with several students from traditional Rwandan universities, many students at other universities will only take an internship in their final (often fourth) year at university as a university requirement, whereas Kepler encourages students to take on internships earlier.

For Kepler students, 44% of the 2013 cohort and 72% of the 2014 cohort internships were with Kepler teaching staff or administration. Most Kepler students' internships are during the school year: 89% of Kepler students' internships compared to 38% of comparison students' internships were during the school year. Kepler students' internships tend to be longer, with most internships lasting longer than two months, whereas most comparison students' internships last less than two months (Figure 3). Kepler and comparison students both felt that their internships were relatively well-organized (average ratings of 5.8 and 5.3 respectively on a scale from one to seven, where one is a low level of organization and seven is high).

**Kepler students receive more feedback on their internships from their employers and universities and find the feedback to be more helpful than comparison students do.**

Kepler students on average report getting feedback from their employer more often than comparison students do, and mostly are satisfied with the amount of feedback they receive. They also report that the feedback they receive from employers is more helpful than comparison students report (average ratings of 6.0 and 4.7 respectively on a scale from one to seven, where one is not helpful at all and seven is extremely helpful).

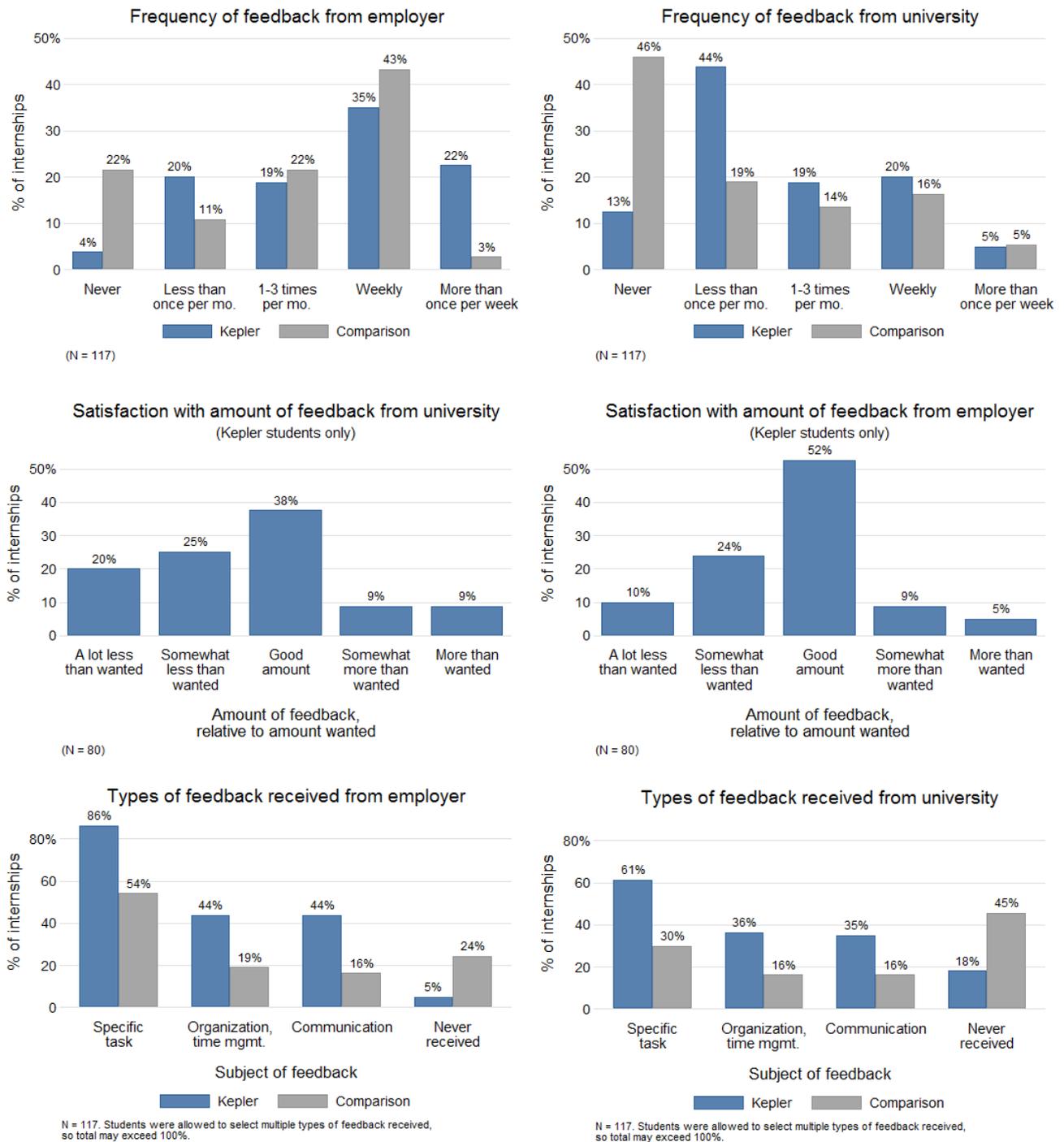
**Both Kepler and comparison students, however, are less satisfied with the amount and quality of feedback on internships that they receive from their universities than feedback from their employers.**<sup>28</sup> Kepler students also receive feedback on their internships from Kepler more frequently than comparison students receive feedback on their internships from their own universities, with 46% of comparison students reporting that they never receive feedback from their universities compared to 13% of Kepler students. However, most Kepler students report receiving feedback either less than once per month or never. Kepler students rate that they find feedback from Kepler to be more helpful than comparison students, though both rate university feedback to be less helpful than employer feedback (average ratings of 5.5 and 4.3 respectively).

Students, both Kepler and comparison, mostly get feedback from their employers and universities on specific tasks, with fewer students receiving feedback on organization, time management, and communication.

---

<sup>28</sup> Kepler regularly checks in with their students' employers to obtain feedback on their students' performance, which they use to shape their curriculum and also give feedback in turn to the students. Comparison students often are required to take internships as part of their degree requirements, typically in their last year, and must submit a report on their internship to their university.

Figure 4: Feedback on internships from employers and universities



Note that only responses from students who reported having at least one internship are included. Students who reported having two internships answered this question for both of their internships, with each answer counting as being from a separate internship.

**Kepler students value communication, time management and organization, and computer skills for internships.** The following results are reported as percent of Kepler students who had internships (counting students with two internships as one student). Students were allowed to list multiple skills, and so percentages add up to more than 100%.

The top skills that Kepler students reported as being helpful for their internships were:

- Communication,<sup>29</sup> including listening and giving feedback (85%)
- Time management, stress management, and other organizational skills (62%)
- Computer skills (61%)

The skills most Kepler students would have liked to have had before their internships were:

- Computer skills, including programming, Microsoft Office applications, photo and video editing, and other software (37%)
- Communication,<sup>30</sup> including listening and giving feedback (30%)
- Other technical skills, such as accounting, marketing, data analysis, and finance (28%)
- Public speaking (25%)
- Time management, stress management, and other organizational skills (25%)

Skills that most Kepler students reported learning in their internships were:

- Communication, including listening, giving feedback, and reporting on work done (48%)
- Time management, planning, and other organizational skills (44%)
- Computer skills (28%)
- Teamwork, collaboration, and other interpersonal soft skills (26%)

## Perceived job prospects

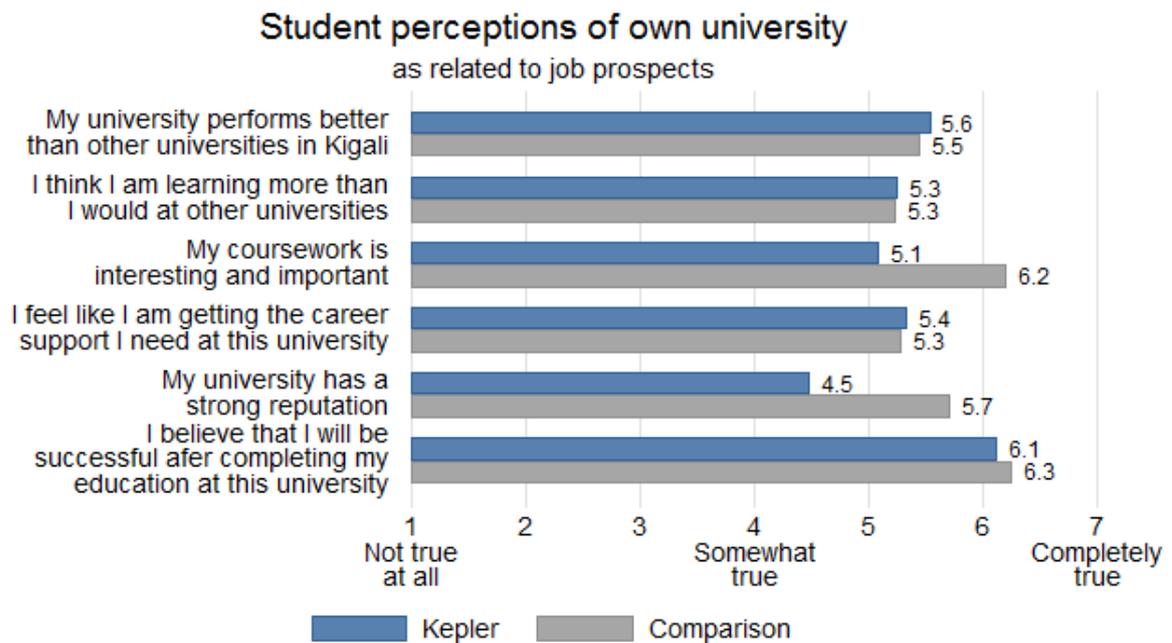
Students were asked several questions that required them to rate how much they agreed with various statements about their university's performance and ability to prepare them for the job market. **Kepler and comparison students on average felt similarly about most aspects of their respective university's performance and ability to prepare students for successful careers, though comparison students more frequently expressed feeling that their coursework is interesting and important, and that their universities have a strong reputation** (Figure 5).

---

<sup>29</sup> Most students simply listed "communication skills," though some specified writing or listening.

<sup>30</sup> Again, most students simply listed "communication skills."

Figure 5



(N = 368)

**Kepler students feel that they can better leverage networks to find jobs with previous employers, university resources, and other personal connections than comparison students do, though there are discrepancies within Kepler students by cohort and gender.** When asked what resources and connections they thought would be the most helpful to them in finding a job, more Kepler than comparison students reported previous employers and university staff as potential resources, as well as active networking, whereas more comparison than Kepler students reported anticipating relying on advertisements (Figure 6).

Figure 7 and Figure 8 break down Kepler students' responses by cohort and by gender, respectively. More of the 2013 cohort than the 2014 cohort reported that previous employers and university staff would be helpful, while more of the 2014 cohort reported that student clubs and advertisements would be helpful. Compared to male Kepler students, slightly fewer Kepler female students report that university, previous employer, and family connections would help, with more reporting instead that other resources, especially active networking, will be more helpful.

Figure 6

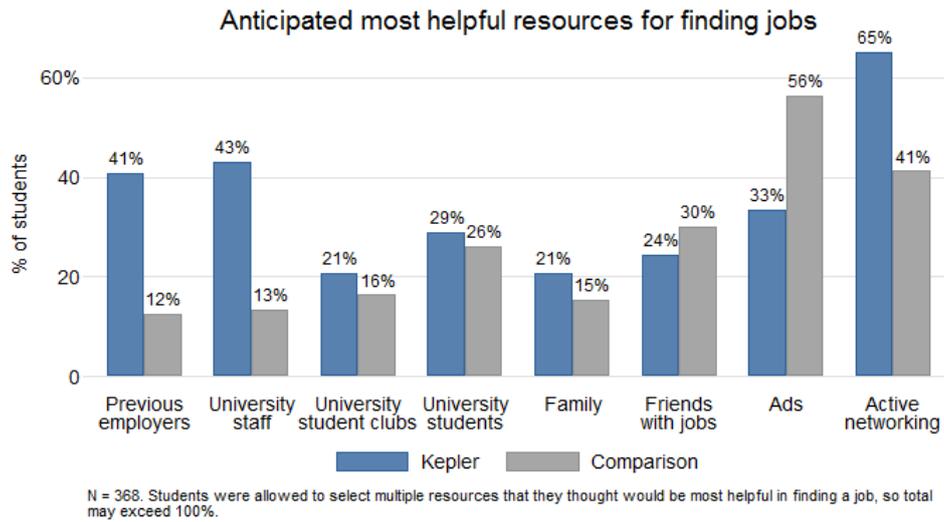


Figure 7

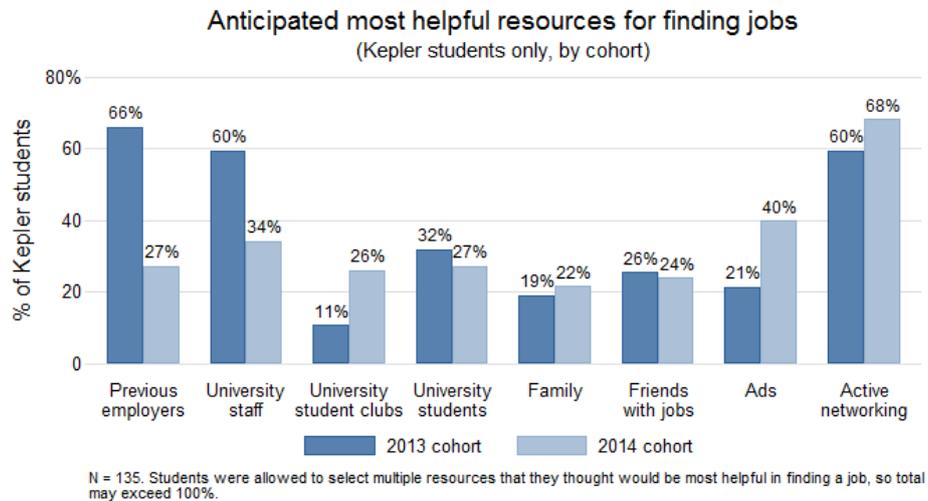
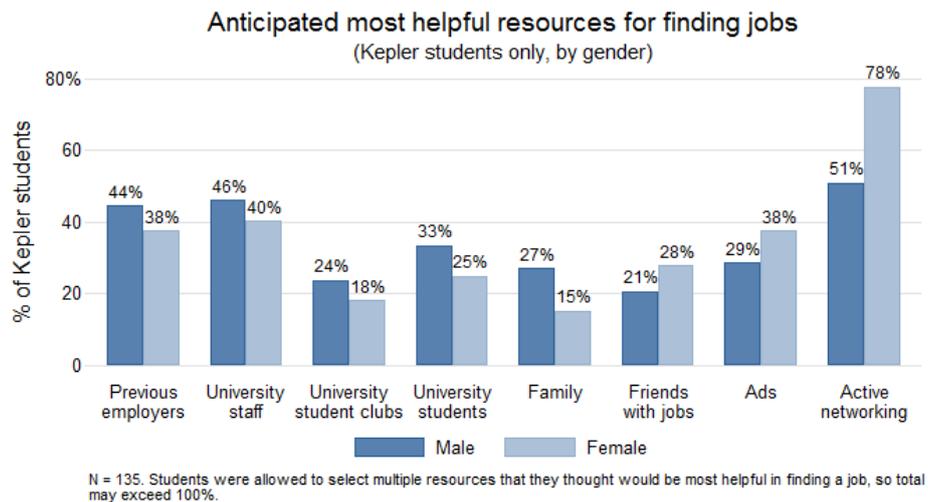
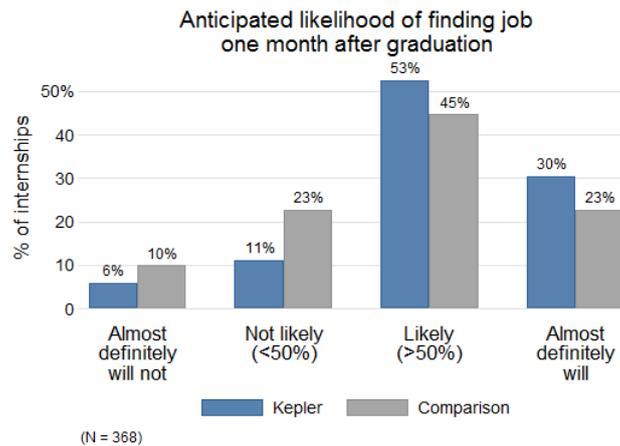


Figure 8



When asked how likely they think they are to find a job within one month after graduation, **Kepler students are on average more confident that they will find a job** (Figure9).

Figure 9



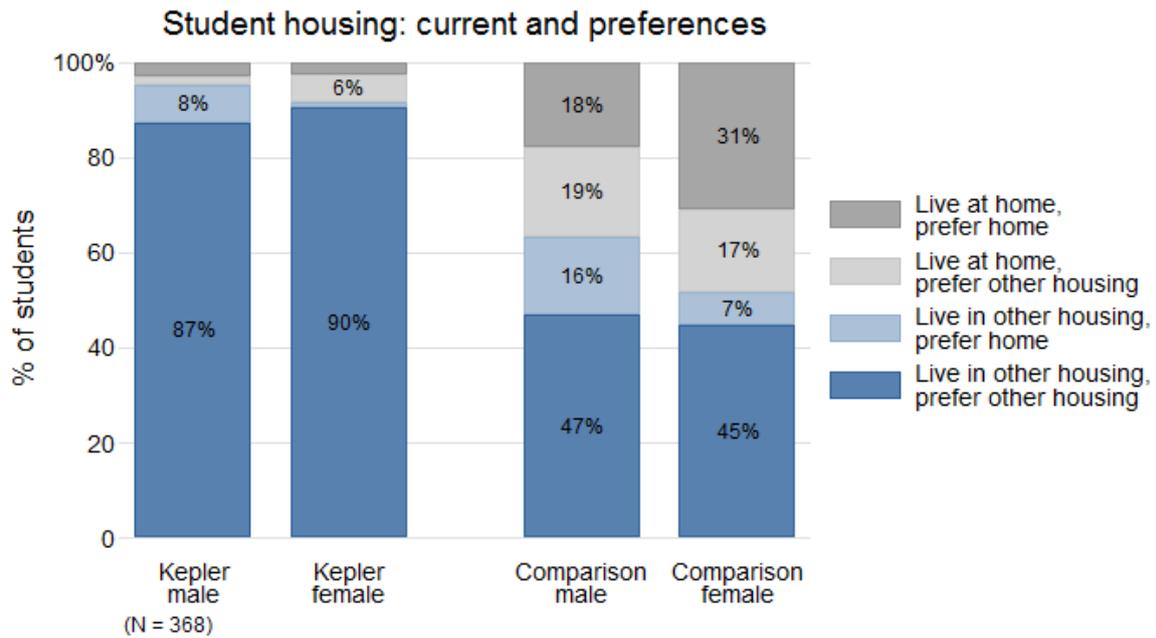
## Housing

While almost all Kepler students live in Kepler student housing, just over half of the comparison students live away from home, mostly in dormitories or other rented, shared housing. Compared to male students, a somewhat greater proportion of female comparison students live at home.

**Almost all Kepler students report that they would find it easier to do school work in student housing rather than at home**, with just a handful of 2014 cohort students saying they would prefer living at home. A majority of comparison students also said doing school work would be easier in other (student) housing. **The top reason for students' housing preference—both for students who preferred to live in other housing and who preferred to live at home—is the ability to get assistance with assignments.**<sup>31</sup> Students with preferences for living outside of home mentioned being able to work in groups or getting assistance from classmates as advantages; students living at home cited advantages getting help from parents as an advantage.

<sup>31</sup> In the multiple choice question asking why students said it would be easier to do school work in their selected location, options given were fewer chores and household responsibilities, fewer distractions from having to talk to other people, fewer distractions from noise, and ability to get assistance with assignments.

Figure 10: Where students live and where they would find it easier to do school work



Students were asked where they normally live while taking university classes, as well as whether they would find it easier to do school work at home or in other housing. Other housing for Kepler students refers to their student housing, and for comparison students mostly refers to dormitories or other rented, shared housing. Bars that represent less than 3% have their labels omitted in this figure.

## Academic dishonesty

**On average, Kepler students perceive a range of academically dishonest activities to be more unethical than comparison students**, with the most unethical activities being turning in the same paper as a friend and copying exam answers (Figure 11).

When asked to rate how much they felt different factors pressured students to cheat (Figure 12), **both Kepler and comparison students feel the most pressing factor is time pressure to finish school work when the student is busy with other things**. Kepler students feel that social pressure from roommates and classmates are the next most pressing factors (to a similar degree), while comparison students feel that competitive pressure to do well is a larger source of pressure.

Figure 11

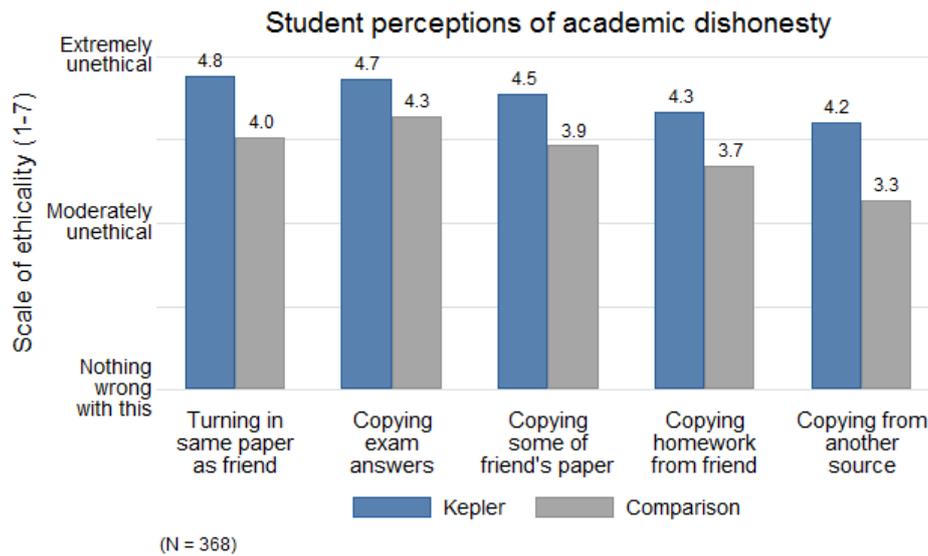
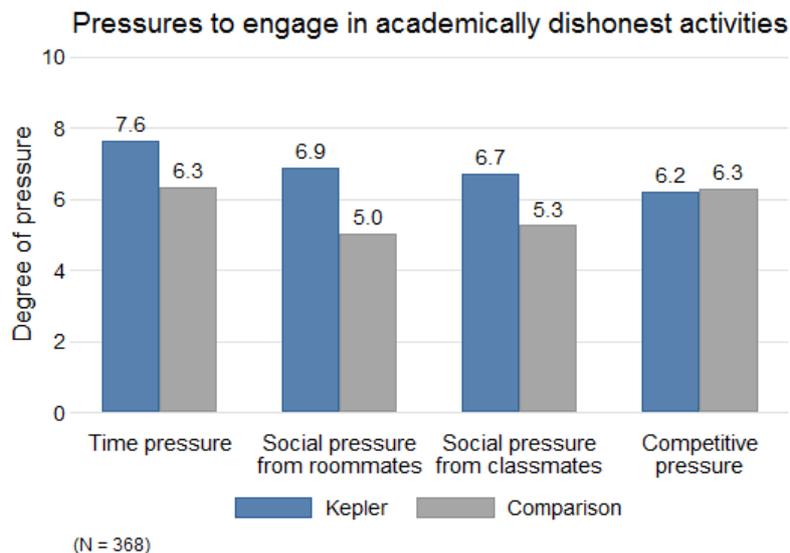


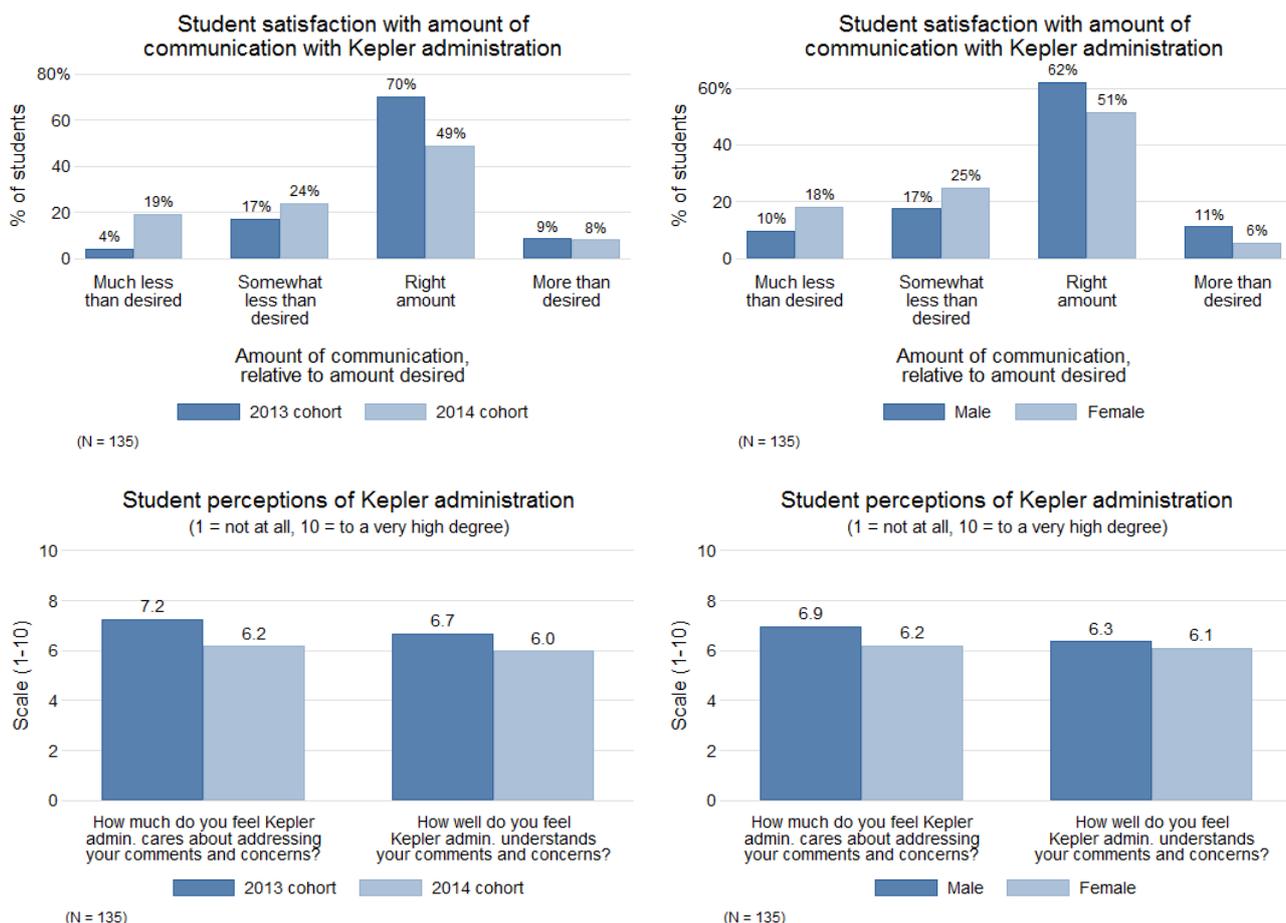
Figure 12: Self-reported relative importance of different factors that may pressure students to engage in academically dishonest activities



## Communication with Kepler administration

Kepler students were asked about their satisfaction with the amount of communication they have with the Kepler administration and about how much they feel Kepler administration cares about and understands their concerns. The figures below present these results broken down by cohort and by gender. **In general, the 2013 cohort and male students are more satisfied with their relationship with Kepler administration compared to the 2014 cohort and female students respectively, both in terms of their satisfaction with the amount of communication they have with the administration and how much their concerns are cared for and understood.**

Figure 13: Kepler student perceptions of communication with the Kepler administration, by cohort and by gender



In an open-ended question asking students for feedback for the Kepler administration on communication, many students responded that they felt positively about the Kepler administration, and thought the administration was doing its best. Other common themes in the responses included:

- Students want to be **involved more in the decision-making process** on issues that affect students directly, particularly in cases of negative perceived consequences.
- Students would like **communication on a regular basis** (e.g. once or twice a month), and more **immediate communication** on issues that affect students.
- Several **students felt that Kepler did not treat and value all students** equally. For example, some felt that some of their peers received preferential treatment in receiving internships and noted that they would like a **more even playing field**.

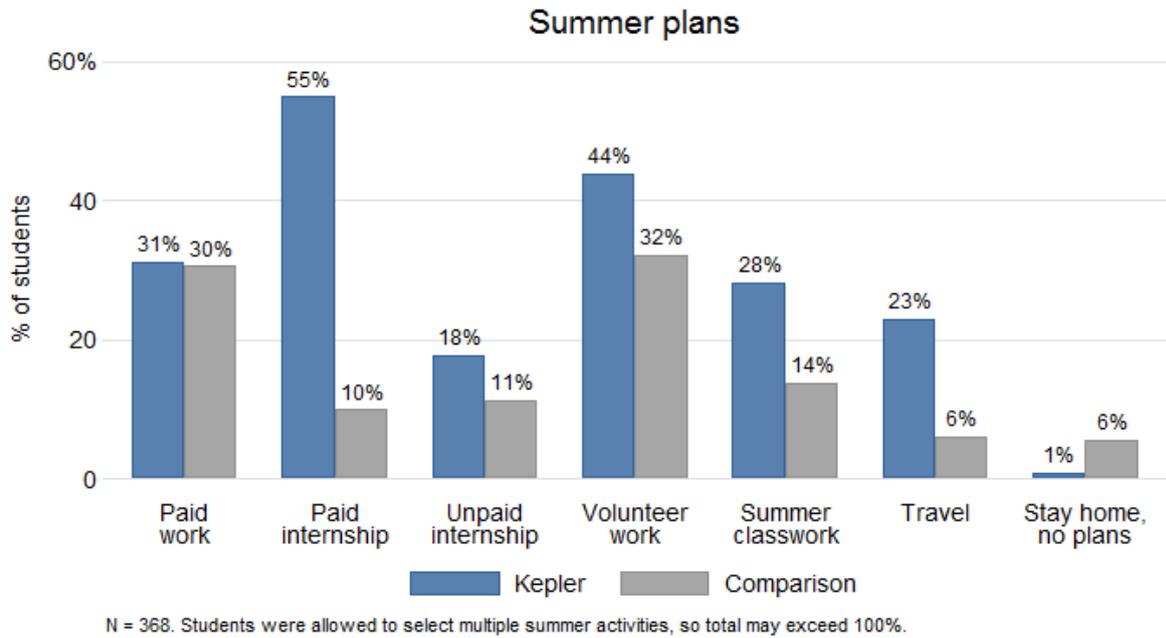
Some students raised other individual concerns related to not feeling respected or fully understood by the Kepler administration.

## Summer plans

Kepler students' summer plans differ on average from the comparison group's (Figure 14), with Kepler students planning to take on more internships (especially paid), volunteer work,

summer classwork,<sup>32</sup> and travel. In general, each Kepler student on average reports engaging in more summer activities than each comparison student.

Figure 14



<sup>32</sup> For most Kepler students, summer classwork involves working on projects for College for America (CfA).

## Conclusions

Results overall suggest that, for the 2014 cohort, the Kepler program has had significant impacts on their students' critical thinking, cognitive skills, English language skills, and computer literacy relative to traditional Rwandan universities. It is unclear, however, whether increases in Kepler students' English skills are in part responsible for the higher critical thinking test scores.

These test results together with the survey findings suggest that Kepler students are being better prepared than their peers in several ways for the job market. With active facilitation by and support from Kepler to pursue internships, more Kepler students are engaging in internships earlier on in their university education, and are receiving more feedback on their work than their peers are who did have internships. While Kepler students perceive their university to be performing mostly at a similar level as comparison students, though less strong in importance of coursework and reputation, they also feel that they have more personal connections they can use in finding jobs and are more confident in their likelihood to find jobs after graduation.

## Future evaluation interests

While this evaluation examines some impacts of the Kepler program on the first two cohorts of Kepler students, further research on learning outcomes of the same cohorts would be valuable in charting the students' progress during their entire Kepler experience.

Further evaluation of the Kepler model as it continues to develop may be helpful to identify the effects of changes to the model by comparing the performance and characteristics of different cohorts. However, as these early evaluations seem to support that the initial Kepler model is having significant impact on several learning outcomes, it may make sense to wait and evaluate a more mature model of the program. A future evaluation that is able to use a randomized controlled design would be even more rigorous than the current design. While Kepler has expressed willingness to randomize admission to the program for a portion of an incoming class, doing so would require more qualified applicants than available spaces in the Kepler program.

In addition to tracking learning outcomes, Kepler has expressed interest in tracking student employment outcomes both before graduation, when many Kepler students take internships and jobs, and after graduation. This is particularly valuable as Kepler shifts its goal more explicitly toward preparing its students to be competitive in the job market. Potential outcomes of interest include professional competencies and soft skills on the job, income, and leadership in the workplace and community. Such an evaluation may involve surveys, soft skills tests, and qualitative interviews with both (graduated) students and employers.

## Appendix A: Technical details on methodology

### Matching method for 2014 cohort<sup>33</sup>

Matching for the 2014 cohort was done using propensity score matching, stratified (grouped) by certain baseline learning outcomes. All 207 comparison students who completed the Year 2 baseline testing were included in the pool of students used to match with the 88 Kepler students in the 2014 cohort<sup>34</sup> in order to form the final 2014 cohort comparison group.

Propensity score matching matches students based on their propensity scores, which are estimated probabilities for each student that he or she would be in the Kepler class, taking into account the characteristics of both the Kepler and comparison students. For the 2014 cohort, a logistic regression with the following key covariates was used to estimate the propensity score:

- Age
- Gender
- Urban/rural domicile (self-identified)
- Poverty level, as estimated by the Progress out of Poverty Index (PPI)
- Senior 5 and Senior 6 marks
- Years of computer use
- Scores on all baseline tests (IELTS reading; IELTS writing; Watson-Glaser critical thinking; English, math, and logic components of the cognitive skills test; aggregate computer literacy test score; and typing speed)

Students were matched using their propensity scores within four strata (groups), defined by whether students performed above or below the Kepler average for the IELTS writing and typing speed tests. These two tests had the greatest difference in scores between the Kepler and unmatched comparison students, and so were stratified on in an effort to reduce bias in these scores. Each Kepler student was matched without replacement with two comparison students, when possible;<sup>35</sup> however, in the higher-scoring strata where there were fewer than two comparison students for every one Kepler student, some Kepler students were only matched with one comparison student.

143 comparison students that best matched the Kepler group in terms of their propensity score probability estimates were invited to participate in the endline testing, with 139 completing the tests.

---

<sup>33</sup> Matching for the 2013 cohort was also done using propensity score matching, though without stratification on baseline learning outcomes as they had no baseline test scores. Please refer to the Year 1 Endline Report for further details on matching for the 2013 cohort.

<sup>34</sup> There were 88 first year Kepler students in the 2014 cohort as of April 2015. This includes one 2014 cohort student who was on academic probation, but likely to rejoin classes. One student who entered Kepler in 2013 but, due to personal circumstances, was moved to the 2014 cohort is still treated as being part of the 2013 cohort for the purposes of this evaluation.

<sup>35</sup> Matching without replacement means that a comparison student could not be matched with multiple Kepler students—that is, they could not be “replaced” into the pool of comparison students to match from after being matched once.

## Limitations to matching methodology

The Year 2 evaluation uses a matching strategy. While matching methods can ensure that the treatment and comparison groups are balanced on key observable characteristics that predict probability of entering treatment and outcomes, there are still limitations to this methodology.

Matching assumes that all relevant background characteristics are measured. However, there are often characteristics such as ambition that cannot be fully quantified or measured. Even if they could be, matching all quantified variables is not always possible, especially with a small sample size. Matching can balance unobserved characteristics between the treatment and control groups insofar as the observable matched characteristics are correlated with unobserved relevant background characteristics, but cannot guarantee that there are no significant differences in unobserved characteristics that could, in addition to treatment status, drive differences in outcomes.

## ANCOVA regression analysis

All treatment effects used to estimate impacts for this evaluation (listed in the Methodology - Design section) were estimated using analysis of covariance (ANCOVA), whereby a regression was run to determine the relationship between treatment and comparison group test results, including important covariate<sup>36</sup> controls. For the 2014 cohort, in estimating the treatment effect for a given test, the Year 2 baseline scores from the test were included as a covariate to adjust for performance on the test prior to attending university. Covariates used in the analysis for this report were:

- Age
- Gender
- Urban/rural domicile (self-identified)
- Poverty level, as estimated by the Progress out of Poverty Index (PPI)
- Whether both parents' of the students were alive
- Senior 5 and Senior 6 marks
- Exposure to English at home: whether English is the primary language spoken at home, or whether the student's mother or father speaks English
- Years of computer use
- Household ownership of computers

## Kepler vs. comparison group balance on covariates

The tables below list the averages and standard deviations<sup>37</sup> of covariates and baseline scores for the Kepler and matched comparison students who were included in the analysis for this report,<sup>38</sup> along with the difference in averages and corresponding  $p$ -values<sup>39</sup> between Kepler and comparison students.

---

<sup>36</sup> Covariates refer to characteristics that are likely to affect the outcome of interest.

<sup>37</sup> Standard deviations are not given for binary variables.

<sup>38</sup> All 2014 cohort Kepler students enrolled in Kepler as of April 2015 completed the Year 2 endline tests. Of the 143 comparison students from the 2014 cohort invited to take the Year 2 endline tests, 139 students completed all the tests. Students who did not complete the tests were unable to either due to personal circumstances or because they had moved abroad for scholarships.

Table 1: Covariates used for analysis and matching

Covariate	Kepler mean (std dev.) (N = 88)	Comp. mean (std dev.) (N = 139)	Difference (p-value)
	Age <sup>ab</sup>	20.9 (2.4)	20.6 (2.1)
Gender (% female) <sup>ab</sup>	54.5%	52.4%	2.1% (p = 0.758)
Domicile (% urban) <sup>ab</sup>	42.0%	44.1%	-2.0% (p = 0.766)
Progress out of Poverty Index (PPI) <sup>ab</sup>	48.4 (10.2)	48.2 (9.7)	0.2 (p = 0.867)
Both parents alive <sup>a</sup>	56.8%	58.0%	-1.2% (p = 0.856)
Exposed to English at home <sup>a</sup>	21.6%	28.2%	-6.6% (p = 0.269)
Senior 5 marks <sup>ab</sup>	75.3 (9.7)	74.8 (8.0)	0.5 (p = 0.645)
Senior 6 marks <sup>ab</sup>	74.6 (9.9)	73.5 (8.1)	1.1 (p = 0.375)
Attended private secondary school	58.0%	36.4%	21.6%*** (p = 0.001)
Years of computer use <sup>ab</sup>	4.8 (2.3)	4.9 (2.8)	-0.1 (p = 0.739)
Household owns computer <sup>a</sup>	22.7%	23.1%	-0.3% (p = 0.951)
Expected earnings post- graduation (RWF)	398352 (234280)	331483 (218482)	66870 RWF** (p = 0.029)
Expected earnings 5 years after graduation (RWF)	1443352 (1500000)	1761455 (2000000)	-318102 RWF (p = 0.195)

<sup>a</sup> used as covariate in ANCOVA analysis of test results for this report

<sup>b</sup> used as covariate in 2014 cohort matching

<sup>39</sup> A *p*-value denotes the statistical significance of the test comparing averages of a variable between Kepler and comparison students. Using a 5% significance level, a *low p*-value (less than or equal to 0.05) means that the two groups were *statistically different* in the average of the variable in question. A *high p*-value (greater than 0.05) means that we cannot conclude that the two groups are *statistically different*. In this report, one asterisk indicates a statistical difference at the 10% significance level, two asterisks indicate a statistical difference at the 5% significance level, and three asterisks represent a statistical difference at the 1% significance level.

Table 2: Baseline test scores for 2014 cohort

<i>Baseline test score</i>	<i>Kepler mean (std dev.) (N = 88)</i>	<i>Comp. mean (std dev.) (N = 139)</i>	<i>Difference (p-value)</i>
<i>Watson-Glaser critical thinking</i>	48.1% (11.9 pp)	48.5% (9.9 pp)	-0.5% (p = 0.733)
<i>Cognitive skills – English</i>	51.3% (22.3 pp)	50.3% (21.8 pp)	0.9% (p = 0.750)
<i>Cognitive skills – Math</i>	40.7% (19.9 pp)	44.5% (18.3 pp)	-3.8% (p = 0.138)
<i>Cognitive skills – Logic</i>	52.7% (19.0 pp)	50.6% (20.0 pp)	2.2% (p = 0.415)
<i>IELTS Reading</i>	33.0% (14.4 pp)	31.6% (12.7 pp)	1.4% (p = 0.425)
<i>IELTS Writing</i>	72.5% (15.5 pp)	70.2% (13.6 pp)	2.3% (p = 0.252)
<i>Computer literacy</i>	26.3% (10.8 pp)	25.0% (11.2 pp)	1.3% (p = 0.391)

Baseline test scores were used as covariates in matching and in ANCOVA analysis of test results for this report for the 2014 cohort. Note that the baseline computer literacy score does not include email.

## Appendix B: Sample test questions

### Critical thinking: Watson-Glaser II Critical Thinking Appraisal

#### Directions:

An inference is a conclusion that a person can draw from certain observed or supposed facts. For example, if the lights are on in a house and music can be heard coming from the house, a person might infer that someone is at home. But this inference may or may not be correct. It is possible that the people of the house did not turn the lights and the radio off when they left the house.

*In this test, each exercise begins with a statement of facts that you are to regard as true. After each statement of facts you will find several possible inferences i.e., conclusions that some persons might draw from the stated facts. Examine each inference separately, and make a decision as to its degree of truth or falsity.*

For each inference you will find spaces in the answer sheet labelled T, PT, ID, PF and F. For each inference put a cross on the answer sheet under the appropriate heading as follows:

**T** if you think the inference is definitely TRUE; that it properly follows beyond a reasonable doubt from the statement of facts given.

**PT** if, in the light of the facts given, you think the inference is PROBABLY TRUE; that it is more likely to be true than false.

**ID** if you decide that there are INSUFFICIENT DATA; that you cannot tell from the facts given whether the inference is likely to be true or false; if the facts provide no basis for judging one way or the other.

**PF** if, in the light of the facts given, you think the inference is PROBABLY FALSE; that it is more likely to be false than true.

**F** if you believe the inference is definitely FALSE; that it is wrong, either because it misinterprets the facts given, or because it contradicts the facts or necessary inferences from those facts.

#### Statement:

**Two hundred school students in their early teens voluntarily attended a recent weekend student conference in Leeds. At this conference, the topics of race relations and means of achieving lasting world peace were discussed, since these were problems that the students selected as being most vital in today's world.**

#### Proposed Inferences (and Answers):

1. As a group, the students who attended this conference showed a keener interest in broad social problems than do most other people in their early teens. (PT, because, as is common knowledge, most people in their early teens do not show so much serious concern with broad social problems. It cannot be considered definitely true from the facts given because these facts do not tell how much concern other young teenagers

may have. It is also possible that some of the students volunteered to attend mainly because they wanted a weekend outing.)

2. The majority of the students had not previously discussed the conference topics in the schools. (PF, because the students' growing awareness of these topics probably stemmed at least in part from discussions with teachers and classmates.)
3. The students came from all parts of the country. (ID, because there is no evidence for this inference.)
4. The students discussed mainly industrial relations problems. (F, because it is given in the statement of facts that the topics of race relations and means of achieving world peace were the problems chosen for discussion.)
5. Some teenage students felt it worthwhile to discuss problems of race relations and ways of achieving world peace. (T, because this inference follows from the given facts; therefore it is true.)

## Cognitive skills

### *English grammar and vocabulary*

1. Identify the adjective in this sentence:  
*Alice ran quickly to avoid the scary predator.*
  - predator
  - avoid
  - quickly
  - scary
2. Select the word the best completes this sentence:  
*The boy was \_\_\_\_\_ when his brother moved away.*
  - hungry
  - special
  - energetic
  - upset

### *Math*

1. At a fruit market, Mary bought 2 avocados and 4 bananas for 800 RWF. If an avocado costs twice as much as a banana, how much does one avocado cost?
  - 100 RWF
  - 200 RWF
  - 400 RWF
  - 600 RWF

2. If Michael can peel and eat one groundnut every 3 seconds, how long would it take him to eat 30 groundnuts?
- 10 seconds
  - 30 seconds
  - One and a half minutes
  - Three minutes

### Logic

1. Assume the *first two* statements are true.
- (1) All actors are wealthy
  - (2) Pleasance is not an actor
  - (3) *Pleasance is not wealthy.*

If the first two statements are true, the third statement is:

- True
  - False
  - Uncertain, not enough information
2. The uncle of the father of my nephew is my:
- Cousin
  - Uncle
  - Father
  - Grandfather

## English language skills: International English Language Testing System (IELTS) reading

Do the following statements agree with the claims of the writer in Reading Passage 1?

**YES** if the statement agrees with the claims of the writer  
**NO** if the statement contradicts the claims of the writer  
**NOT GIVEN** if it is impossible to say what the writer thinks about this

1. Helium is a very cold substance.
- Yes
  - No
  - Not given

Complete the summary below.

Choose **NO MORE THAN TWO WORDS** from the passage for each answer.

Write your answers below.

Sobotka argues that big business and users of helium need to help look after helium stocks because 2 ..... will not be encouraged through buying and selling alone.

## English language skills: International English Language Testing System (IELTS) writing

Write about the following topic:

**These days, a person's worth seems to be judged according to social status and material possessions. Old-fashioned values, such as honour, kindness and trust, no longer seem important.**

To what extent do you agree or disagree with this opinion?

Give reasons for your answer and include any relevant examples from your own knowledge or experience. Write at least 250 words.

## Computer literacy

### *Typing speed*

Students were instructed to complete one of the three-minute typing tests at [www.typingtest.com](http://www.typingtest.com). Adjusted typing speed in words per minute (wpm) was calculated number of characters typed divided by five to get number of words typed, subtracting off the number of errors made as a penalty, and then divided by three to get words per minute.

### *Web research*

Use the internet to find answers to the following questions. Write the answers in the spaces below. You will be graded on the accuracy of your answer.

1. Who finished third in the Women's race of the Boston Marathon in 2013?

### *Website credibility*

Determine if the sites would be considered a **credible** website.

**Credible** means that it is likely to have reliable information that you could cite in an academic paper.

Yahoo Answers (<https://answers.yahoo.com/question/index?qid=20150413224540AAhQped>)  
YES | NO

CIA World Fact Book (<https://www.cia.gov/library/publications/the-world-factbook/geos/rw.html>)

YES | NO

### *Microsoft Word*

Students were instructed to open a pre-prepared Microsoft Word document and make edits such as the following:

- Change the font (letter style) of the document to **Times New Roman**.
- Add **10-pt spacing** between new lines.
- Insert a **footnote reference** at the end of the title that says “Release Date: April 23, 2011.”

### *Microsoft Excel*

Students were instructed to open a pre-prepared Microsoft Excel spreadsheet and make edits such as the following:

- **Insert a row** at the top of table with words “Uganda Internet Data Plans” in the first cell.
- Add a column with numbers next to the price column. In this column divide the price by **2,500** using a formula.
- **Make a scatterplot** with columns **MB** and **price**.

### *Email*

Students were given the following instructions to send an email to a given email address:

Write a formal email, stating that you are submitting the computer test documents, attaching the Word and Excel documents.

Students without an email account were given a Gmail email account and password to use.

## Appendix C: Year 2 endline raw test score statistics

The following table lists the means and standard deviations for the tests administered at the Year 2 endline in April-May 2015. These statistics have not been adjusted for differences in covariate values, i.e. they do not account for differences in scores that may be due to student demographics or baseline scores (unlike the test results presented in the main results section of this report).

Table 3: Year 2 endline (April-May 2015) test scores for 2014 cohort

<i>Test</i>	<i>Kepler</i>			<i>Comparison</i>		
	<i>N</i>	<i>Mean</i>	<i>Std. Dev.</i>	<i>N</i>	<i>Mean</i>	<i>Std. Dev.</i>
<i>Watson-Glaser critical thinking</i>	88	<b>54.5%</b>	8.2 pp	139	<b>49.7%</b>	9.4 pp
<i>Cognitive skills - English</i>	88	<b>43.2%</b>	19.2 pp	139	<b>34.2%</b>	20.0 pp
<i>Cognitive skills - Math</i>	88	<b>41.4%</b>	17.9 pp	139	<b>40.9%</b>	17.5 pp
<i>Cognitive skills - Logic</i>	88	<b>42.4%</b>	19.1 pp	139	<b>34.4%</b>	19.2 pp
<i>IELTS reading</i>	88	<b>36.0%</b>	17.0 pp	139	<b>29.7%</b>	15.1 pp
<i>IELTS writing</i>	88	<b>56.1%</b>	5.9 pp	139	<b>53.2%</b>	4.7 pp
<i>Computer literacy</i>	88	<b>60.0%</b>	11.0 pp	139	<b>33.5%</b>	13.9 pp
<i>Typing speed (wpm)</i>	88	<b>26.9 wpm</b>	8.1 wpm	139	<b>10.0 wpm</b>	5.8 wpm

Standard deviations for all test scores (except for typing speed) are listed in percentage points (pp). Typing speed is given in words per minute (wpm). The computer literacy test consists of typing speed, web research, website credibility, Microsoft Word, Microsoft Excel, and email components.

## Appendix D: Regression results

Below are tables that list coefficients (with p-values in brackets) from regression results for the primary analysis. All test scores (except for the typing speed score, noted below) range from 0 to 1. In all tables, one asterisk indicates statistical significance at the 10% significance level, two asterisks indicate statistical significance at the 5% significance level, and three asterisks represent statistical significance at the 1% significance level.

Please refer to Appendix A for further details on covariates used.

Regression table 1: First year impact for 2014 cohort

Variable	Watson- Glaser	Cog. Sk. English	Cog. Sk. Math	Cog. Sk. Logic	IELTS reading	IELTS writing	Computer literacy
Treatment	0.525*** [0.000]	0.452*** [0.000]	0.144 [0.221]	0.414*** [0.002]	0.387*** [0.002]	0.596*** [0.000]	1.900*** [0.000]
Baseline score	0.166*** [0.007]	0.296*** [0.000]	0.494*** [0.000]	0.259*** [0.000]	0.433*** [0.000]	0.311*** [0.000]	0.437*** [0.000]
Age (years)	-0.047 [0.146]	0.004 [0.925]	-0.048* [0.067]	-0.068** [0.017]	-0.123*** [0.001]	-0.067** [0.031]	-0.092*** [0.000]
Female (binary)	0.034 [0.813]	-0.218 [0.128]	-0.241** [0.040]	-0.099 [0.471]	-0.122 [0.330]	0.101 [0.518]	-0.445*** [0.000]
Exposure to English at home (binary)	-0.030 [0.850]	-0.011 [0.947]	0.059 [0.707]	0.242 [0.117]	-0.053 [0.725]	0.073 [0.707]	-0.053 [0.668]
Urban (binary)	-0.020 [0.892]	0.272* [0.077]	0.015 [0.907]	0.335** [0.041]	0.270** [0.041]	0.108 [0.523]	0.060 [0.599]
PPI score	0.003 [0.649]	-0.014** [0.032]	0.002 [0.774]	0.009 [0.252]	0.003 [0.655]	-0.008 [0.284]	-0.009* [0.067]
Both parents alive (binary)	0.044 [0.737]	0.143 [0.286]	0.068 [0.581]	0.167 [0.215]	0.301** [0.011]	0.134 [0.357]	-0.066 [0.506]
Senior 5 marks	0.001 [0.932]	0.004 [0.798]	0.012 [0.273]	0.015 [0.261]	-0.001 [0.941]	0.028 [0.110]	-0.002 [0.884]
Senior 6 marks	0.010 [0.523]	0.005 [0.724]	-0.001 [0.905]	-0.004 [0.765]	0.016 [0.169]	-0.021 [0.223]	-0.007 [0.550]
Years of computer use	0.017 [0.537]	0.013 [0.623]	0.020 [0.457]	-0.009 [0.725]	0.007 [0.751]	-0.022 [0.459]	0.041** [0.042]
Household owns computers (binary)	0.101 [0.589]	0.155 [0.420]	-0.130 [0.406]	-0.357* [0.052]	-0.059 [0.708]	-0.019 [0.929]	0.192 [0.144]
Constant	-0.104 [0.921]	-0.248 [0.825]	0.111 [0.914]	0.070 [0.946]	1.041 [0.309]	1.178 [0.345]	2.963*** [0.000]
Observations	227	227	227	227	227	227	227
R-squared	0.140	0.210	0.329	0.198	0.423	0.194	0.733

The dependent variable in these regressions was the April-May 2015 test score for the indicated test.

## Appendix E: Robustness checks on the effect of English skills on critical thinking test scores

Because all tests were administered in English, and Kepler students show greater improvements in their English skills than comparison students overall, it is possible that improvements in English skills could be driving increases in scores for tests that are meant to assess other skills. This is particularly a concern for the Watson-Glaser critical thinking test, which relies on careful reading and interpretation of short texts in English. To look into this effect, robustness checks were performed for the critical thinking results for the first year impact for the 2014 cohort.

### Methodology

The 2014 cohort had several different metrics of English ability: IELTS reading score, IELTS writing score, and score on the English grammar and vocabulary portion of the cognitive skills test. The IELTS reading score and, in a separate analysis, the grammar and vocabulary score, were selected as the most relevant proxies for English ability that would be helpful in taking the Watson-Glaser critical thinking test.

Using these metrics, several additional variables were added to the ANCOVA regression used in the analysis for critical thinking, where the Year 2 endline Watson-Glaser score (expressed in standard deviations of comparison students' scores) is the dependent variable:

- **Year 2 baseline English score:** The coefficient on this term reflects how much baseline English ability influences endline critical thinking scores. The 2014 cohort Kepler and comparison students had on average similar English test scores at baseline (see Table 2).
- **Interaction of Year 2 baseline English score with treatment status:** The coefficient on this interaction term represents how much Kepler's effect on critical thinking scores operates through baseline English ability. A positive coefficient would suggest that higher baseline English ability is more helpful in increasing critical thinking scores for Kepler students than for comparison students, possibly because English forms a more central part of the Kepler curriculum than at comparison universities.
- **Change in English score from Year 2 baseline to endline:** The coefficient here represents how much improvements in English ability drive endline critical thinking scores. Because Kepler students generally experienced greater improvements in English test scores than comparison students, a positive coefficient would suggest that this discrepancy in English improvement may be responsible for Kepler students' higher critical thinking scores at endline.
- **Interaction of change in English score with treatment status:** This coefficient reflects how much Kepler's effect on critical thinking scores operates through increasing English ability. A positive coefficient would suggest that improved English ability is more helpful in increasing critical thinking scores for Kepler students than for comparison students, again possibly because English is more necessary to learn well at Kepler than at comparison universities.

For these added variables, the baseline and endline English scores are included as the percentage of total possible points on the respective test, ranging in value from 0 to 1. With the inclusion of these additional variables, the coefficient on treatment status alone reflects the “pure” effect of the first year of Kepler education on Watson-Glaser critical thinking scores that does not relate to English ability.

## Results

When using scores from the English grammar and vocabulary part of the cognitive skills test, the “pure” effect of Kepler on critical thinking scores unrelated to English abilities is attenuated (0.41 SD, as opposed to 0.53 SD) and no longer statistically significant. Coefficients on both baseline grammar and vocabulary scores and change in scores are positive, while coefficients on both of these terms interacted with treatment status are very small and not statistically significant. This suggests that stronger English abilities, as captured by the grammar and vocabulary test, accounts for at least some of the Kepler students’ increase in Watson-Glaser scores.

However, results using the IELTS reading scores are less intuitive. The “pure” effect of Kepler on critical thinking scores is actually higher in magnitude (0.69 SD), but less statistically significant. While baseline IELTS reading scores seem to be an important predictor of endline critical thinking scores, as evidenced by a large and positive coefficient, improvement in reading scores does not seem to be.

Overall, these results suggest that at least part of the increase in Watson-Glaser critical thinking scores for Kepler students can likely be attributed to stronger English skills, though the evidence is not entirely clear.

Regression table 2: Robustness checks on the effect of English skills on Watson-Glaser critical thinking scores

<i>Variable</i>	<i>Original analysis</i>	<i>With grammar and vocab. scores</i>	<i>With IELTS reading scores</i>
<i>Treatment</i>	0.525*** [0.000]	0.406 [0.271]	0.687* [0.067]
<i>English baseline score</i>	-	1.257** [0.018]	2.608*** [0.001]
<i>English baseline score * Treatment</i>	-	0.040 [0.958]	-0.780 [0.482]
<i>Change in English score</i>	-	0.917** [0.035]	0.560 [0.397]
<i>Change in English score * Treatment</i>	-	-0.060 [0.929]	0.731 [0.449]
<i>Watson-Glaser Year 2 baseline score</i>	0.166*** [0.007]	0.108* [0.086]	0.089 [0.146]
<i>Age (years)</i>	-0.047 [0.146]	-0.040 [0.174]	-0.024 [0.455]
<i>Female (binary)</i>	0.034 [0.813]	0.086 [0.532]	0.022 [0.871]
<i>Exposure to English at home (binary)</i>	-0.030 [0.850]	-0.063 [0.693]	-0.077 [0.613]
<i>Urban (binary)</i>	-0.020 [0.892]	-0.101 [0.488]	-0.041 [0.776]
<i>PPI score</i>	0.003 [0.649]	0.005 [0.423]	0.003 [0.696]
<i>Both parents alive (binary)</i>	0.044 [0.737]	-0.004 [0.973]	-0.005 [0.972]
<i>Senior 5 marks</i>	0.001 [0.932]	0.002 [0.901]	0.005 [0.747]
<i>Senior 6 marks</i>	0.010 [0.523]	0.007 [0.631]	-0.001 [0.971]
<i>Years of computer use</i>	0.017 [0.537]	0.012 [0.661]	0.005 [0.866]
<i>Household owns computers (binary)</i>	0.101 [0.589]	0.111 [0.554]	0.064 [0.711]
<i>Constant</i>	-0.104 [0.921]	4.124*** [0.000]	4.083*** [0.000]
<i>Observations</i>	227	227	227
<i>R-squared</i>	0.140	0.179	0.209

The dependent variable in these regressions is the Year 2 endline Watson-Glaser score, expressed in terms of standard deviations of the comparison students' test scores.

## Appendix F: Subgroup results

The figures and tables below give the results for subgroup analyses by gender and urban/rural domicile. These analyses examine whether Kepler had a greater impact on female or male Kepler students relative to female and male comparison students respectively, and similarly for rural and urban students. For gender, where male students generally score higher than female students, this can be thought of as looking at whether Kepler is narrowing or widening the gap between male and female students, relative to the gender gap that exists between male and female comparison students.

The subgroup analysis is done using difference-in-differences estimation, where (using gender as an example) the difference between male and female Kepler students' scores is subtracted from the difference between male and female comparison students' scores.

Very few results are statistically significant at the 5% significance level. These subgroup analyses only have sufficient statistical power<sup>40</sup> to detect very large differences between subgroups. As such, particularly for differences which are rather large but not statistically significant, we cannot necessarily rule out the possibility of there being a meaningfully large difference between subgroups.

However, while a few results are statistically significant, caution should be taken in placing much weight on them. Because many subgroup analyses were performed, it is possible that some results will turn out to be statistically significant by random chance. Any results of interest which are statistically significant would be worth following up on in subsequent evaluations to see whether the results are repeated, which would strengthen confidence in the result.

---

<sup>40</sup> Statistical power is the probability that a study will detect an effect when there is truly an effect present. Statistical power can be increased by having a greater number of observations (here, students) in each group that is being compared. The subgroup analyses presented here in general have less statistical power than the main analyses because more groups are being compared, but there is still the same total number of observations.

## Gender

Both among Kepler and comparison students, males outperformed females on nearly all Year 2 endline exams. This gap is likely due to pre-existing differences in male and female students' performance, and was also found in Kepler's own admissions data. In the results below, a positive difference indicates that there was a greater impact for female students (that is, Kepler narrowed the gender gap), while a negative difference indicates a greater impact for male students (widening the gender gap). Results from subgroup analyses do not provide much evidence for Kepler either narrowing or widening the pre-existing gender gap by having a greater or lesser impact on students by gender.

Table 4: Subgroup results by gender

<i>Tests</i>	<i>Std. effect size</i>	<i>Perc. pt. diff.-in-diff.</i>	<i>p-value</i>
<i>Watson-Glaser critical thinking</i>	-0.16 SD	-1.5 pp	0.536
<i>Cognitive skills - English</i>	0.15 SD	3.0 pp	0.554
<i>Cognitive skills - Math</i>	-0.30 SD	-5.2 pp	0.235
<i>Cognitive skills - Logic</i>	-0.10 SD	-1.9 pp	0.708
<i>IELTS reading</i>	0.13 SD	2.0 pp	0.588
<i>IELTS writing</i>	0.12 SD	0.5 pp	0.716
<i>Computer literacy</i>	0.12 SD	1.7 pp	0.527

## Urban/rural domicile

Endline scores did not differ systematically across all tests for students from urban and rural backgrounds for either Kepler or comparison students. In all results below, a positive difference indicates that there was a greater impact for rural students while a negative difference indicates a greater impact for urban students. Overall, the results from the subgroup analyses do not provide much evidence for Kepler having a greater impact on either urban or rural students.

Table 5: Subgroup analysis by urban/rural domicile

<i>Tests</i>	<i>Std. effect size</i>	<i>Perc. pt. diff.-in-diff.</i>	<i>p-value</i>
<i>Watson-Glaser critical thinking</i>	-0.00 SD	-0.0 pp	0.994
<i>Cognitive skills - English</i>	0.13 SD	2.6 pp	0.604
<i>Cognitive skills - Math</i>	-0.12 SD	-2.1 pp	0.606
<i>Cognitive skills - Logic</i>	0.22 SD	4.3 pp	0.401
<i>IELTS reading</i>	0.44 SD*	6.6 pp	0.069
<i>IELTS writing</i>	-0.68 SD**	-3.2 pp	0.031
<i>Computer literacy</i>	0.42 SD**	5.9 pp	0.022